Chapter 1

Introduction

1.1 Motivation

Linear models have been very useful and popular in applied medical research. These models are simple to implement and they have the additional advantage of ease of interpretation. However, for more complex applied problems lesser known generalizations of linear models are needed. Within the biomedical sciences there are many data sets of underlying complex processes that cannot be handled by means of linear regression. For example, individuals are measured across time or across different items inducing correlation or some biomedical treatment may have one optimal value resulting in a nonlinear response. For more complex data sets, non-linear methods are more powerful tools in terms of unbiasedness as they provide clear insight of nonlinear relationship between response and predictors. The aim of this thesis is to describe explanatory and predictive models that are generalizations of linear regression.

1.2 Linear Models

Linear models have been very useful and popular as statistical methods for the analysis of medical data. They are frequently used in medical research where it is often the case what one wants to predict the outcome of a disease (Y) through the covariates $(X_1, \ldots, X_p)$. In linear models, the relationship between a response and the predictors in the data set is described in a linear manner. Linear models are simple in terms of implementation and inference.
A standard linear model has the form

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + \epsilon, \tag{1.1}$$

where

- $X_1, \ldots, X_p$ are the independent variables also called the predictors or covariates. These covariates can be original variables in the data, their transformations, combinations, polynomials or interactions among the covariates [12].
- $\beta_0, \beta_1, \ldots, \beta_p$ are the unknown parameters which are estimated using the data set and are also called regression coefficients. β_0 is the intercept and β_j is the coefficient that determines the contribution of the covariate X_j.
- ϵ is known as a random error term and assumed to be independently and identically distributed with constant variance, $\epsilon \sim \mathcal{N}(0, \sigma^2)$. This term basically represents the difference between the original and observed values of Y in the model.
- Y is a response variable which is assumed to have a normal distribution, $Y \sim \mathcal{N}(\mu, \sigma^2)$. The response variables is also called the dependent variable or the outcome variable and predicted through the covariates. The mean is given by $\mu + \sum_{i=1}^{p} \beta_i (x_i)$.

Given a random sample of size n observations $(Y_i, X_{i1}, \ldots, X_{ip})$, $i = 1, \ldots, n$, the linear model (1.1) can be written in matrix representation form as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.2}$$

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1p} \\ 1 & X_{21} & X_{22} & \ldots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \ldots & X_{np} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \ \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

The key element in the linear model is the estimation of the parameters, $\boldsymbol{\beta}$. There are several methods that can used for parameters estimation, for example, maximum likelihood estimation or the method of least squares [34], among them. We describe here the maximum likelihood estimation for $\boldsymbol{\beta}$ estimation in the linear model. Given the distribution of $Y_i \sim N(\mu_i, \sigma^2)$ where $\mu_i = \boldsymbol{X_i'}\boldsymbol{\beta} = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}$,

the probability density function (PDF) of Y is

$$f(Y_i;\boldsymbol{\beta},\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(Y_i-\mu_i)^2\right), \tag{1.3}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(Y_i-\boldsymbol{X}_i'\boldsymbol{\beta})^2\right). \tag{1.4}$$

The likelihood function is given by

$$L\left(\boldsymbol{\beta},\sigma^2\right) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(Y_i-\boldsymbol{X}_i'\boldsymbol{\beta})^2\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)\right), \tag{1.5}$$

and the log-likelihood is

$$\log L\left(\boldsymbol{\beta},\sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} - \frac{1}{2\sigma^2}\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right). \tag{1.6}$$

The maximum likelihood estimation of $\boldsymbol{\beta}$ is obtained by differentiating w.r.t. to β and equating to zero as follows:

$$\begin{aligned}
\frac{\partial \log L\left(\boldsymbol{\beta},\sigma^2\right)}{\partial\boldsymbol{\beta}} &= 0\\
\frac{1}{2\sigma^2}\frac{\partial}{\partial\boldsymbol{\beta}}\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right) &= 0\\
\frac{\partial}{\partial\boldsymbol{\beta}}\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right) &= 0\\
-2\boldsymbol{X}'\left(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}\right) &= 0\\
\boldsymbol{X}'\boldsymbol{Y} &= \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}.
\end{aligned}$$

It follows that the maximum likelihood estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{1.7}$$

An estimate of the variance of $\widehat{\boldsymbol{\beta}}$ is given by

$$\text{var}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}, \tag{1.8}$$

similarly, the maximum likelihood estimation gives an estimate of σ^2 in the form of

$$\widetilde{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(Y_i - \boldsymbol{X_i'}\widehat{\boldsymbol{\beta}}\right)^2}{n}. \tag{1.9}$$

However, the usual unbiased estimator of σ^2, which differs from this maximum likelihood estimator, is given by

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(Y_i - \boldsymbol{X_i'}\widehat{\boldsymbol{\beta}}\right)^2}{n-(p+1)}. \tag{1.10}$$

Note that the difference between $\widetilde{\sigma}^2$ and $\widehat{\sigma}^2$ vanishes as $n \to \infty$.

1.3 Generalized Linear Models

In the linear model described in the previous section, the response variable is continuous and assumed to follow a normal distribution. However, the linear model is not appropriate for binary data, counts, and categorical response variables because the underlying distributional assumptions of the model do not hold. The generalized linear model (GLM) are an extension of linear model that allow modelling a non-normal response as well as non-linear relationship. GLM were first introduced as statistical modeling tool in 1972 by Nelder and Wedderburn [31].

A generalized linear model (GLM) consists of three components:

1. A random component, specifying the conditional distribution of the response variable Y_i for the ith of n independently sampled observations, given the values of the explanatory variables in the model. In the GLM formulation the distribution of Y_i is a member of an exponential family. The general form of the exponential family of distribution for Y has a probability density function (PDF) that can be written as

$$f_\theta\left(y\right) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c\left(y,\phi\right)\right). \tag{1.11}$$

   In the above equation a, b and c are arbitrary functions, θ is the canonical parameter of the exponential family distribution and ϕ is the dispersion parameter [11, 37]. The probability distributions

of the exponential family contains several well -known distributions for example, binomial, poisson, normal and gamma distributions, depending on the circumstances [37].

2. A linear predictor η_i that is a linear function of the covariates,

$$\eta_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip}.$$

3. A smooth and invertible linearizing link function $g(\cdot)$, which transforms the expectation of the response variable, $\mu_i = E(Y_i)$, to the linear predictor:

$$g(\mu_i) = \eta_i.$$

When $\theta = \eta$ in (1.11), then the link function is called canonical link function.

Consider n independent observations $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ and assume that they belong to the same exponential family. Then the likelihood function using the general exponential family of distribution given in (1.11) is

$$L(\boldsymbol{\theta}) = \exp\left(\sum_{i=1}^{n} \frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c\left(Y_i, \phi\right)\right). \tag{1.12}$$

The log-likelihood is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c\left(Y_i, \phi\right), \tag{1.13}$$

$$= \sum_{i=1}^{n} \ell_i(\theta_i), \tag{1.14}$$

where $g(\mu_i) = \eta_i$, $\eta_i = \boldsymbol{X}'\boldsymbol{\beta}$, $\mu_i = E(Y_i) = \partial b(\theta_i)/\partial\theta_i$, and

$$\ell_i(\theta) = \frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c\left(Y_i, \phi\right),$$

depending on $b(\theta)$ and $c\left(Y, \phi\right)$ we have different distributions. For example, the Bernoulli distribution with parameter π can be expressed in the form of exponential family of distributions for generalized linear models as follows. The probability mass function of a Bernuolli

distribution is

$$\begin{aligned} f(Y;\pi) &= \pi^Y(1-\pi)^{1-Y} \\ &= \exp\left(\log\left[\pi^Y(1-\pi)^{1-Y}\right]\right) \\ &= \exp\left(Y\log\pi + 1 - Y\log(1-\pi)\right) \\ &= \exp\left(Y\left[\log\pi + \log(1-\pi)\right] + \log(1-\pi)\right) \\ &= \exp\left(Y\log\frac{\pi}{1-\pi} + \log(1-\pi)\right) \\ f(Y;\theta) &= \exp\left(Y\theta - b(\theta)\right), \end{aligned}$$

such that $\theta = \log\frac{\pi}{1-\pi}$, $b(\theta) = \log\frac{e^\theta}{1+e^\theta}$, and $c(Y,\phi) = 1$. Moreover, since $E[Y] = \pi$ and having $\theta = \eta$ we have the logit link function $g(\pi) = \text{logit}(\pi) = \log\frac{\pi}{1-\pi}$. Note that, the mean and variance of a generalized linear model is given by

$$\begin{aligned} \mu_i =& E(Y_i) = \frac{\partial b(\theta_i)}{\partial\theta_i}, \\ \text{var}(Y_i) =& a(\phi)\frac{\partial^2 b(\theta_i)}{\partial{\theta_i}^2}. \end{aligned}$$

In medical sciences, a commonly used model is the binomial distribution combined with the logit link function, also known as logistic regression, where the outcome of a treatment Y is binary,e.g. yes=1 and no=0. In the case when the response variable has the binomial distribution then the GLM takes the form

$$g\left[E\left(Y_i\right)\right] = \text{logit}\left(p\left(\boldsymbol{X}_i\right)\right) = \log\frac{p\left(\boldsymbol{X}_i\right)}{1-p\left(\boldsymbol{X}_i\right)} = \boldsymbol{X}_i^{'}\boldsymbol{\beta}, \tag{1.15}$$

where $p\left(\boldsymbol{X}_i\right) = Pr[Y_i = 1 \mid \boldsymbol{X}_i = \boldsymbol{x}_i]$, $E\left(Y_i\right) = \mu_i$ and g is the logit link function. The link function basically connects the mean of the response $E\left(Y_i\right)$ to the linear predictors $\boldsymbol{X}_i^{'}\boldsymbol{\beta}$ of the model. Different exponential families have different link functions, for example, logistic regression has logit link function, Poisson regression has log link function. The properties of GLM have been discussed in detail by McCullagh and Nelder [28].

1.3.1 Maximum likelihood estimation (MLE)

Maximum likelihood estimation for generalized linear models provides estimates of the regression parameters and their estimated asymp-

totic standard errors.

The maximum likelihood estimate of $\boldsymbol{\beta}$ is the value obtained by maximizing the log-likelihood in (1.13). We first take the first partial derivatives with respect to β_j as

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{\partial \ell_i(\theta_i)}{\partial \beta_j}, \\ &= \sum_{i=1}^{n} \frac{\partial \ell_i(\theta_i)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \ell_i(\theta_i)}{\partial \mu_i} &= \frac{\partial \ell_i(\theta_i)}{\partial \theta_i} \Big/ \frac{\partial \mu_i}{\partial \theta_i} = \frac{Y_i - \frac{\partial b(\theta_i)}{\partial \theta_i}}{a(\phi)} \Big/ \frac{\partial^2 b(\theta_i)}{\partial {\theta_i}^2} = \frac{Y_i - \mu_i}{\mathrm{Var}(Y_i)} \\ \frac{\partial \eta_i}{\partial \beta_j} &= X_{ij}.\end{aligned}$$

Then we have

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{Y_i - \mu_i}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} X_{ij}. \tag{1.16}$$

In principle, the maximum likelihood estimator can be obtained by equating the score function to zero and solving for β_j; that is,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{Y_i - \mu_i}{\mathrm{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} X_{ij} = 0, \quad j = 1, \ldots, p+1. \tag{1.17}$$

However, solving these score equation does not always provide closed form solutions. Consequently, numeric methods are used to find the estimates that maximize the log-likelihood function. The Fisher scoring approach is proposed as a general method for the numerical evaluation of $\widehat{\boldsymbol{\beta}}$ in generalized linear models. The method of scoring is an iterative approach where at the mth iteration, the updated estimates are given by

$$\widehat{\boldsymbol{\beta}}^{(m)} = \widehat{\boldsymbol{\beta}}^{(m-1)} + \left[\mathcal{I}^{(m-1)}\right]^{-1} \mathcal{U}^{(m-1)}, \tag{1.18}$$

where $\mathcal{U}$ is the score vector given in (1.16) and $\mathcal{I}$ is the Fisher infor-

mation matrix with the (j,k)th entry is defined by

$$\mathcal{I}_{jk} = -E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial\beta_j \partial\beta_k}\right] \tag{1.19}$$

$$= E\left[\frac{\partial \ell(\boldsymbol{\theta})}{\partial\beta_j}\frac{\partial \ell(\boldsymbol{\theta})}{\partial\beta_k}\right] \tag{1.20}$$

$$= E\left\{\left[\sum_{i=1}^{n}\frac{Y_i-\mu_i}{\text{Var}(Y_i)}\frac{\partial\mu_i}{\partial\eta_i}X_{ij}\right]\left[\sum_{l=1}^{n}\frac{Y_l-\mu_l}{\text{Var}(Y_l)}\frac{\partial\mu_l}{\partial\eta_l}X_{ik}\right]\right\}. \tag{1.21}$$

This simplifies to

$$\mathcal{I}_{jk} = \sum_{i=1}^{n}\frac{X_{ij}X_{ik}}{\text{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2. \tag{1.22}$$

Let W be an $(n\times n)$ diagonal matrix with

$$w_{ii} = \frac{1}{\text{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2.$$

Then, the Fisher information matrix can be rewritten as

$$\mathcal{I} = \boldsymbol{X}'W\boldsymbol{X}.$$

Multiplying both sides of equation (1.18) by $\mathcal{I}^{(m-1)}$, we get

$$\mathcal{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m)} = \mathcal{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} + \mathcal{U}^{(m-1)}. \tag{1.23}$$

The right hand side of equation (1.23) is a vector with the jth element given by

$$\begin{aligned}\left(\mathcal{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} + \mathcal{U}^{(m-1)}\right)_j &= \sum_{k=0}^{p}\sum_{i=1}^{n}\frac{X_{ij}X_{ik}}{\text{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2\widehat{\beta}_k^{(m-1)} \\ &\quad + \sum_{i=1}^{n}\frac{Y_i-\mu_i}{\text{Var}(Y_i)}\frac{\partial\mu_i}{\partial\eta_i}X_{ij} \\ &= \sum_{i=1}^{n}X_{ij}\frac{1}{\text{Var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\sum_{k=0}^{p}\Bigg(X_{ik}\widehat{\beta}_k^{(m-1)} \\ &\quad + (Y_i-\mu_i)\left(\frac{\partial\mu_i}{\partial\eta_i}\right)\Bigg),\end{aligned}$$

so that the right hand side of equation (1.23) is given by

$$\mathcal{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} + \mathcal{U}^{(m-1)} = \boldsymbol{X}^{'} W \boldsymbol{z}, \tag{1.24}$$

where W is evaluated at $\widehat{\boldsymbol{\beta}}^{(m-1)}$ and the vector $\boldsymbol{z}$ has elements

$$z_i = \sum_{k=0}^{p} X_{ik}\widehat{\beta}_k^{(m-1)} + (Y_i - \mu_i)\left(\frac{\partial \mu_i}{\partial \eta_i}\right).$$

Similarly, the left hand side of equation (1.23) can be expressed as

$$\mathcal{I}^{(m-1)}\widehat{\boldsymbol{\beta}}^{(m-1)} = \boldsymbol{X}^{'} W \boldsymbol{X} \widehat{\boldsymbol{\beta}}^{(m-1)}. \tag{1.25}$$

Finally, substituting (1.24) and (1.25) into (1.23), the Fisher scoring iterative method can be written as

$$\boldsymbol{X}^{'} W \boldsymbol{X} \widehat{\boldsymbol{\beta}}^{(m-1)} = \boldsymbol{X}^{'} W \boldsymbol{z}.$$

This has the same form as normal equations for a linear model based on weighted least squares. However, this has to be solved iteratively. In such a case the maximum likelihood estimators are obtained by the method of iterative weighted least squares. The standard errors for the estimates $\widehat{\boldsymbol{\beta}}$ can be calculated as the square roots of the diagonal elements of

$$\text{Cov}\left(\widehat{\boldsymbol{\beta}}\right) = a(\phi)\left(\boldsymbol{X}^{'} W \boldsymbol{X}\right)^{-1}, \tag{1.26}$$

which is obtained from the final iterative weighted least squares iteration. If $a(\phi)$ in the above equation is unknown, then estimate is required.

1.4 Linear Mixed Models (LMM)

Clustering or grouped data structures commonly arise in applied medical statistics. Sometimes clustering structure of data set is complex, for example in the case of longitudinal data or repeated measures [24], where the standard statistical independence assumption is unwarranted. Linear mixed effects model is a statistical modeling tool that is suitable for handling clustered or grouped data. A mixed effects model consist on two effects, the fixed effects and the random effects [11, 32].

The fixed effects are the unknown coefficients which are to be estimated [11]. The random effects are random variables [11] where the statistical units or subjects are drawn at random from a finite population and the focus is to estimate the parameters that describes the distribution of random effects on each subject. Pinheiro and Bates provides full details of linear mixed modelling [32], as well as the implementation of linear mixed models in lme4 package of R [2].

1.4.1 Model

A general linear mixed model (LMM) has the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{1.27}$$

where

- $\boldsymbol{y}$ is a response variable of model.
- $\boldsymbol{X}$ is a $n \times p$ design matrix of fixed effects.
- $\boldsymbol{\beta}$ is a unknown vector $\in \Re^p$ of fixed effects associated with design matrix X.
- $\boldsymbol{Z}$ is a design matrix of $n \times q$ random effects.
- $\boldsymbol{\gamma}$ is unknown vector $\in \Re^q$ of random effects associated with design matrix Z.
- the random effects of $\boldsymbol{\gamma}$ are assumed as normally distributed with mean 0 and positive definite covariance matrix $\boldsymbol{\Sigma}$.
- the noise ϵ is assumed to be normally distributed with mean 0 and covariance matrix $\boldsymbol{R}$.

Then, the variance y given the random effect is

$$\boldsymbol{v}ar\left(\boldsymbol{Y}|z\right) = \boldsymbol{R}, \tag{1.28}$$

such that

$$\boldsymbol{v}ar\left(\boldsymbol{Y}\right) = \boldsymbol{Z\Sigma Z}^t + \boldsymbol{R}.$$

Then, the distribution of response variable is assumed to follow

$$Y \sim \mathcal{N}(\boldsymbol{X\beta}, \boldsymbol{Z\Sigma Z}^t + \boldsymbol{R}). \tag{1.29}$$

The representation in (1.29) tells us that the fixed effects $\boldsymbol{X\beta}$ contributes to the mean of Y, whereas the random effect $\boldsymbol{Z}$ and the noise variance contribute to the variance of $\boldsymbol{Y}$.

1.4.2 Within-subjects and between subjects correlation

Chapter 2 of this thesis focuses on within and between subjects correlation estimates of positive and negative affects of psychiatry patients in terms of a linear mixed model. We introduce the idea with the help of a simple example. Lets assume we have a random intercept model

$$y_{ijt} = \gamma^w_{ijt} + \gamma^b_{jt}, \tag{1.30}$$

where y_{ijt} is a continuous response variable which describes positive affect (PA) or negative affect (NA) of psychiatry patients, where i is a person/patient (subject) $i = 1, \ldots, n$ and j is a PA or NA item and t is time. The γ^w_{ijt} represents within-subject random effect and γ^b_{jt} represent between-subjects random effects. The modeling of random effects usually works with the assumption of a normal distribution. The random effect γ^w_{ij} and γ^b_j are assumed to be distributed as bivariate normal, that is

$$\gamma^b_t = \begin{bmatrix} \gamma_{PA,t} \\ \eta_{NA,t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \boldsymbol{\rho_B}\sigma_1\sigma_2 \\ \boldsymbol{\rho_B}\sigma_1\sigma_2^2 & \sigma_2^2 \end{bmatrix}\right), \tag{1.31}$$

$$\gamma^w_{it} = \begin{bmatrix} \gamma_{i,PA,t} \\ \gamma_{i,NA,t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \boldsymbol{\rho_W}\tau_1\tau_2 \\ \boldsymbol{\rho_W}\tau_1\tau_2^2 & \tau_2^2 \end{bmatrix}\right), \tag{1.32}$$

where

- ρ_B is within-subject correlation and σ_1^2 and σ_2^2 are variance of $\gamma_{PA,t}$ and $\gamma_{NA,t}$ respectively.

- ρ_W is between -subjects correlation τ_1^2 and τ_2^2 of are variance $\gamma_{i,PA,t}$ and $\gamma_{i,NA,t}$, respectively.

We can write the between-subjects and within-subjects model above in a general matrix form, from (1.26), where we assume no fixed effect part of the mixed model, $X = 0$. The matrix representation of this can be shown as below

$$\boldsymbol{Y} = \begin{bmatrix} Y_{111} \\ \vdots \\ Y_{11T} \\ Y_{121} \\ \vdots \\ Y_{12T} \\ \vdots \\ Y_{n11} \\ Y_{11T} \\ Y_{n21} \\ Y_{12T} \end{bmatrix}, \boldsymbol{Z}_{2nT\times(2nT+2n)} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_{PA} \\ \gamma_{NA} \\ \gamma_{1,PA} \\ \gamma_{1,NA} \\ \vdots \\ \gamma_{n,PA} \\ \gamma_{n,NA} \end{bmatrix}.$$

That is our model formulation in (1.30) can be expressed as a linear mixed model, form as in equation (1.26) without the fixed effect part:

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}. \tag{1.33}$$

Thus, the general theory of linear mixed effects model can be applied to our model formulation. In the section that follows we discuss maximum likelihood estimation of parameters in a linear mixed effects model.

1.4.3 Estimation: Restricted Maximum Likelihood Estimation (REML)

Maximum likelihood estimation:
The marginal likelihood for linear mixed effects model from (1.29) can be expressed as

$$L(\beta, V) = (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})V^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\right). \tag{1.34}$$

The parameters β and V are unknown but fixed that can be estimated from the data. Then maximum likelihood estimation based on the marginal log likelihood resulting from (1.34) is given by

$$l(\beta, V) = \frac{-n}{2}\log(2\pi) - \frac{1}{2}\left[\det(\boldsymbol{V})\right] - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^t \boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{1.35}$$

where $\boldsymbol{V} = \boldsymbol{Z}\boldsymbol{\Sigma}\boldsymbol{Z}^t + \boldsymbol{R}$ is the variance of Y. The maximum likelihood estimation of $\boldsymbol{\beta}$ can be obtained by maximizing the log-likelihood (1.35). The resulting estimator is

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{Y}\right). \tag{1.36}$$

Similarly, one can obtain the maximum likelihood estimators of R and Σ or parameters $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ that define $\boldsymbol{R}$ and $\boldsymbol{\Sigma}$. So that an estimator of V is given by $\widehat{\boldsymbol{V}} = \boldsymbol{Z}\widehat{\boldsymbol{\Sigma}}\boldsymbol{Z}^t + \widehat{\boldsymbol{R}}$. Moreover, the best predictor for $\boldsymbol{\gamma}$ can be obtained from the posterior distribution of $\boldsymbol{\gamma}$ given in data $\boldsymbol{Y}$. That is, the posterior distribution can be derived using Bayes' theorem and has the form

$$\boldsymbol{\gamma} \mid \boldsymbol{Y} \sim \mathcal{N}\left(R\boldsymbol{Z}^t\boldsymbol{V}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \left(\boldsymbol{Z}^t\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} + R^{-1}\right)^{-1}\right),$$

such that the best linear unbiased predictor $(BLUP)$ is given by

$$\widehat{\boldsymbol{\gamma}} = E[\boldsymbol{\gamma} \mid \boldsymbol{Y}] = \widehat{\boldsymbol{R}}\boldsymbol{Z}^t\widehat{\boldsymbol{V}}^{-1}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right).$$

Restricted maximum likelihood estimation:
The direct maximum likelihood estimation of the variance parameters presented above results in under-estimation of the variance due to not accounting for the number of estimated fixed parameters [37]. Alternatively, the restricted maximum likelihood estimation (REML)

method is preferred for estimating the variance parameters associated with a linear mixed model. The REML method measures the fit of the variance parameters by taking the average of the likelihood over all the possible values of β [37]. The likelihood for the REML approach is defined as

$$\boldsymbol{L_{RE}}(\boldsymbol{\theta}) = \int \boldsymbol{L}(\boldsymbol{\beta}, \boldsymbol{V})\, d\boldsymbol{\beta}, \tag{1.37}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{\rho}, \boldsymbol{\tau}, \boldsymbol{\sigma^2}\right)$ are correlations and variance parameters used to define $\boldsymbol{\Sigma}$ and $\boldsymbol{R}$ and then $\boldsymbol{V}$, for example see eq (1.30) and (1.31). Substituting the likelihood of (1.34) into the expression for $\boldsymbol{L}_{RE}$ in (1.37) and integrating out $\boldsymbol{\beta}$, it has been shown that the log-likelihood for the REML approach simplifies to

$$\begin{aligned} \boldsymbol{l_{RE}}(\boldsymbol{\theta}) =& \frac{p-n}{2}\log(2\pi) \\ & -\frac{1}{2}\log\left[\det(\boldsymbol{V})\right] - \frac{1}{2}\operatorname{logdet}\left(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}\right) \\ & -\frac{1}{2}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^t \boldsymbol{V}^{-1}\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right). \end{aligned} \tag{1.38}$$

The variance component estimates are obtained by using numerical optimization algorithms on the log-likelihood in (1.38).

1.5 Generalized Additive Models (GAM)

The generalized additive model (GAM) is an important development of statistical modelling techniques, which was originally proposed by Hastie and Tibishrani in 1986 [19]. A generalized additive model is an extension of the generalized linear model (GLM), in which the relationship between the response and the predictors is described in a nonlinear manner, by means of smooth functions of the predictors. These smooth functions are estimated simultaneously [23] to jointly predict the response variable Y. These smooth functions are linear combinations of known basis functions. Therefore it is called a generalized additive model.

1.5.1 Model

A generalized additive model can written as

$$g\left(E\left[Y_i|x_i\right]\right) = \beta_0 + \sum_{j=1}^{p} f_j\left(x_{ij}\right), \tag{1.39}$$

where Y is the response variable which belongs to some exponential family distribution, $g\left(.\right)$ is the link function, which connects the mean of the response to the linear predictors of the additive model, $g\left(u_i\right) = \eta_i$ where

$$\eta_i = \beta_0 + f_1\left(x_{i1}\right) + f_2\left(x_{i2}\right) + \cdots + f_p\left(x_{ip}\right). \tag{1.40}$$

The functions f_j are the smooth functions. These non-parametric smooth functions are data dependent. The smooth functions show the individual contribution of each of the variables in the GAM model, possibly including low level interactions of smooth function (i.e, bivariate smooths) [21, 23]. The smooth functions of the covariates $f_j\left(x\right)$ are assumed to have the following form

$$f_j\left(x\right) = \sum_{k=1}^{K} \beta_{jk} b_k\left(x\right), \tag{1.41}$$

where b_k are the known basis function (known) and β_{jk} are the unknown coefficients. For example, smooth function f_j can be defined using a cubic polynomial basis $b_1\left(x\right) = 1,\ b_2\left(x\right) = x$, $b_3\left(x\right) = x^2,\ b_4\left(x\right) = x^3$, resulting in

$$f_j(x) = \beta_{j1} + \beta_{j2} b_2(x) + \beta_{j3} b_3(x)^2 + \beta_{jk} b_3(x)^3 .$$

In general, we use a polynomial spline basis [37, 40]. A cubic spline of order 10 is given by

$$\begin{aligned} f(x) = \beta_0 b_0(x) + & \beta_1 b_1(x) + \beta_2 b_2(x) + \beta_3 b_3(x) \\ & + \sum_{u=1}^{6} \beta_{3+u} b_{3+u}\left((x - t_u)_+\right) \end{aligned}$$

where the cubic polynomial basis is given relative to 6 arbitrary (but usually equally spaced on the range of x) knot points $\{t_u \mid u = 1, \ldots, 6\}$,

$$\begin{aligned} b_0(x) &= 1 \\ b_1(x) &= x \\ b_2(x) &= x^2 \\ b_3(x) &= x^3 \\ b_{3+u}\left((x - t_u)_+\right) &= (x - t_u)_+^3 \end{aligned} \tag{1.42}$$

where

$$(x - t_u)_+ = \begin{cases} (x - t_u) & \text{if } x > t_u, \\ 0 & \text{otherwise.} \end{cases}$$

When the link function g is the identity link function and Y is a normally distributed response, then we obtain a linear model, i.e,

$$E[Y_i | x_i] = \beta_0 + \sum_{j=1}^{p} \sum_{k=1}^{K} \beta_{jk} b_k(x_{ij}), \tag{1.43}$$

as we have discussed in section 1.1. It turns out that in that case we can write the GAM in the standard form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.44}$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & b_1^1(x_{11}) & \dots & b_K^1(x_{11}) & \dots & b_1^p(x_{ip}) & \dots & b_K^p(x_{ip}) \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 1 & b_1^1(x_{n1}) & \dots & b_K^1(x_{n1}) & \dots & b_1^p(x_{np}) & \dots & b_K^p(x_{np}) \end{bmatrix},$$

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{15} \\ \vdots \\ \beta_{pk} \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

If Y come from an arbitrary exponential family distribution and the link function is given as in (1.43) then the model can be described as a GLM as described in section 1.2.

1.5.2 Estimation

The smooth function $f_j(x_{ij})$ can be estimated in GAM through the scatter plot smoother [20], and there are different methods available for the estimation of scatter plot smoothing. The log likelihood for model (1.44) for a Gaussian response with an identity link is given by

$$l(\beta) = \log L\left(\boldsymbol{\beta}, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} - \frac{1}{2\sigma^2}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right), \tag{1.45}$$

Estimation of (1.45) could follow approaches described in section 1.2 and 1.3. However, the column dimension of X can be very large, leading to poor estimates of β and very wiggly estimates of the underlying functions f_j, $j = 1, 2, \dots, p$. Therefore, we introduce a penalized approach

$$l_\lambda(\beta) = l(\beta) + \sum_{j=1}^{p} \lambda_j \int_{L_J}^{U_j} \left(f_j''(x)\right)^2 dx. \tag{1.46}$$

The first term $l(\beta)$ in the penalized approach is the likelihood. The second term $\lambda_j \int_0^1 \left(f_j''(x)\right)^2 dx$ is the penalty term where (L_j, U_j) is the observed interval of values for the j^{th} covariate. The parameter λ is a non-negative tuning parameter [21], the controls the amount of smoothness by penalizing the curvature of f_j. The penalty term

basically enforces smoothness by integrating over the curvature of the function. The penalty term can be written as a quadratic form in the parameters,

$$\begin{aligned}
\int_{L_j}^{U_j} \left(f_j^{''}(x)\right)^2 dx &= \int_{L_j}^{U_j} \left(\sum_{k=1}^{K} \beta_{jk} b_k^{''}(x)\right)^2 dx. \\
&= \int_{L_j}^{U_j} \left(\sum_{k=1}^{K} \beta_{jk} b_k^{''}(x)\right) \left(\sum_{l=1}^{K} \beta_{jl} b_l^{''}(x)\right) dx. \\
&= \int_{L_j}^{U_j} \sum_{k=1}^{K} \sum_{l=1}^{K} \beta_{jk}\beta_{jl} b_k^{''}(x) b_l^{''}(x) dx. \\
&= \sum_{k=1}^{K} \sum_{l=1}^{K} \beta_{jk}\beta_{jl} \underbrace{\int_{L_j}^{U_j} b_k^{''}(x) b_l^{''}(x) dx}_{S_{kl}^{(j)}} \\
&= \beta^T S^{(j)} \beta. \qquad (1.47)
\end{aligned}$$

so,

$$\sum_{j=1}^{p} \int_{L_j}^{U_j} \left[f_j^{''}(x)\right]^2 dx = \boldsymbol{\beta^t S^{(j)} \beta}, \qquad (1.48)$$

where S is the known matrix

$$\boldsymbol{S} = \begin{bmatrix} S^{(1)} & 0 & \dots & 0 \\ 0 & S^{(2)} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & S^{(p)} \end{bmatrix}. \qquad (1.49)$$

We consider minimizing the penalized log-likelihood

$$l_\lambda(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \lambda \boldsymbol{\beta^t S \beta}, \qquad (1.50)$$

First, we considering the Gaussian case in (1.45) and ignoring constants with respect to β, the penalized log-likelihood can be written

as

$$\begin{aligned} l_\lambda(\beta) &= (y_i - X\beta)^t (y_i - X\beta) + \lambda\beta^t S\beta \\ &= y^t y - \beta^t X^t y + \beta^t X^t X\beta + \lambda\beta^t S\beta \\ &= C - 2\beta^t X^t y + \beta^t \left(X^t X + \lambda S\right)\beta. \end{aligned} \tag{1.51}$$

Taking derivative with respect to β and setting them equal to zero

$$\begin{aligned} \frac{\partial l\beta}{\partial\beta} &= 2X^t (y - X\beta) + \lambda P(\beta) \\ X^t y - X^t X\beta - \lambda S\beta &= 0 \\ \left(X^t X + \lambda S\right)\beta &= X^t y, \end{aligned} \tag{1.52}$$

so, we find

$$\widehat{\beta} = \left(X^t X + \lambda S\right)^{-1} X^t y. \tag{1.53}$$

The functions $f_j(x)$ minimize the penalized least squares in (1.46) if they are natural cubic splines [23].

1.5.3 Choice of smoothing parameter, λ: Generalized cross validation (GCV)

The value of λ controls the degree of smoothness of the estimator $\widehat{f}$ to the data. If λ is small the data will be under-smoothed and if λ is large the data will be over smoothed [37]. In these cases the estimator the estimator $\widehat{f}$ will not be close to the true function f. A data adaptive choice of λ can be obtained by minimizing

$$\sum_{i=1}^{n} \left(f(x_i) - \widehat{f}(x_i)\right)^2. \tag{1.54}$$

However, minimizing of this criterion is not straightforward due to the unknown true function f. A reasonable choice of λ that minimizes the above criterion can be obtained by using leave-out-one cross validation, that is minimizing

$$CV(\lambda) = \sum_{i=1}^{n} (y_i - \widehat{y}_{\lambda,-i})^2, \tag{1.55}$$

where $\widehat{y}_{\lambda,-i}$ is the predicted value at location i if we did not include i in to original predictor, that is

$$\begin{aligned}\widehat{y}_{\lambda,-i} &= x_i^t \widehat{\beta}_{\lambda,-i} \\ &= x_i^t \left(X_{-i}^t X_{-i} + \lambda S\right)^{-1} X_{-i}^t y_{-i},\end{aligned}$$

where $X_{-i} =$ is matrix X minus row i and y_{-i} is y minus ith observation.
It is computationally inefficient to calculate $CV\left(\lambda\right)$ for each of the n data points by leaving out one data point at a time. Cross validation can be expressed in an efficient way by

$$CV\left(\lambda\right) = \sum_{i=1}^{n} \left(\frac{y_i - \widehat{y}_{\lambda,i}}{H_{ii}}\right)^2 \tag{1.56}$$

where $\widehat{y}_{\lambda,i}$ is the predicted value at location i using the entire n data points, and H is given by

$$H = I - X^t \left(X^t X + \lambda S\right)^{-1} X^t.$$

Furthermore a good approximation of $CV\left(\lambda\right)$ is given by a generalized cross-validation (GCV) criterion

$$GCV\left(\lambda\right) = \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_{\lambda,i}\right)^2}{\sum_{i=1}^{n} H_{ii}}. \tag{1.57}$$

We choose λ that minimizes the GCV criterion.

1.6 Other linear models

Linear models are the basis of many statistical models, where the response variable is continuous and assumed to be normally distributed. The response variable is predicted through the predictors which are linear in the parameters. Generalized linear models are the extension of simple linear models, where the response variable is binary, a count or something that follows some exponential family distribution, for example, binomial, Poisson and Gamma probability distributions. Furthermore, the response variable in a GLM is connected to the linear predictors through a link function g. Eeach exponential family has its own set of on different link functions. Details of LM and GLM have

been discussed in detail in section 1.2 and 1.3 of the thesis. A linear

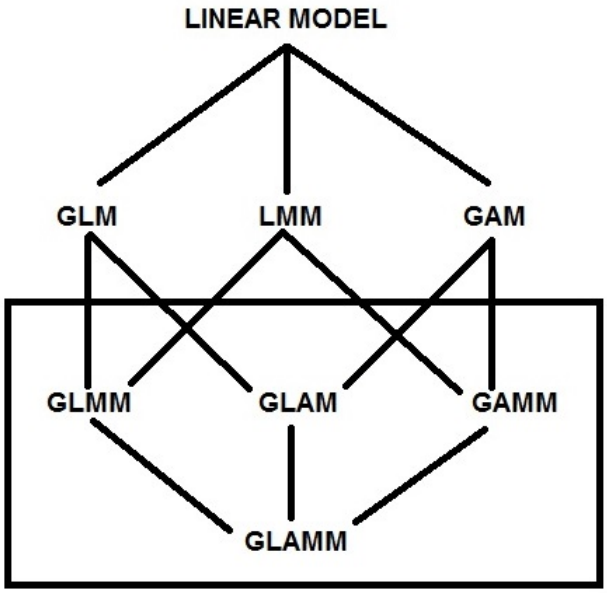

Figure 1.1: Generalizations of linear modelling

mixed model (LMM) is another extensions of a linear model. These models are particularly suitable to handle clustered and longitudinal data. The mixed effect models consist of two effects: the random effects and the fixed effects. Details of LMM have been discussed in section 1.4. Generalized additive models are another extension of simple linear models [12], which deal with non-linear effects of each covariate to produce more biased predictions. The important difference between LM and GAM is the change from a linear relationship between the covariates and response of the model to a nonlinear relationship, where the linear part $\sum_{j=1}^{p} \beta_j x_{ij}$ is replaced by the sum of smooth functions $\sum_{j=1}^{p} f_j\left(x_{ij}\right)$ [20, 23].
Generalized linear mixed model (GLMM) are the extension of LMM and GLM. The GLMMs were introduced by Breslow and Clayton [4] where the response variable follows exponential family distribution and the model contains both fixed and random effects,

$$g(E(y_i)) = X\beta + \boldsymbol{Z}\gamma, \gamma \sim (0, \Sigma)\,,$$

where $g(.)$ is a monotonic link function [38] and Σ is the covariance matrix of the random effects γ.
The generalized additive mixed model (GAMM) is an extension of the LMM and GAM. The main strength of GAMMs is that both smooth fixed effects and random effects are included [26],

$$E(Y_i) = \boldsymbol{Z_i^t}\gamma + \sum_{j=1}^{p} f_j\left(x_{ij}\right).$$

The generalized linear additive models (GLAM) are the extensions of GLM and GAM. The important difference between GAM and GLAM is the response variable. The response variable has an exponential family distribution along with a non linear part in the model,

$$g\left(E\left[Y_i\right]\right) = \beta_0 + \sum_{j=1}^{p} f_j\left(x_{ij}\right). \tag{1.58}$$

Generalized linear additive mixed model (GLAMM) are extensions of GLAM and GAM. In GLAMM the response variable belongs to some exponential family distribution. The general form of GLAMM takes the form

$$g(E(Y_i)) = \boldsymbol{Z}\gamma + \sum_{j=1}^{p} f_j\left(x_{ij}\right),$$

which include a random effect Zr and where the covariates have a nonlinear effect on $g(EY_i)$ and f_j are some smooth functions that are estimated as natural splines $w.r.t$ to some functional basis.

1.7 Contribution of this thesis

The contribution of this thesis consists of the application of, in particular, the linear mixed effects models (Chapter 2) and the generalized additive models (Chapter 3,4). We also developed some methodology for generalized additive modelling of ordinary differential equations (Chapter 5). The data analysis was performed on real biomedical data sets. The contribution of this thesis has been divided into the following four chapters.

Chapter 2: In this chapter, we analyzed longitudinal data of individuals positive affect (PA) and negative affect (NA) using a linear mixed effect model to find the unbiased estimates of within-subjects and between-subjects correlations. PA and NA are two different mood dimensions, which change over time in individuals. A within-subject correlation is not necessarily the same as a between-subject correlation. Our study distinguishes within-subject and between-subject correlations by means of a random effects approach. The result of this study has been submitted for publication.

Chapter 3: In this chapter, we analyzed kidney stone data using a functional regression method namely the generalized additive model. The idea is new in kidney stone analysis for retrograde intra renal surgery (RIRS) treatment. The contribution of this study is that we developed a recommendation chart for RIRS patients based on a generalized additive model. In addition, we investigated predictive parameters that influence the stone free rates, sepsis rate and operation time of the RIRS treatment. This study will be of help to urologists to identify which types of kidney stones patients can benefit from RIRS treatment and what would be their probability of success and of sepsis. This chapter has been submitted for publication.

Chapter 4: This chapter contains a comparative study between Percutaneous nephrolithotomy (PCNL) and retrograde renal surgery (RIRS). Traditionally for larger stone (> 2) cm PCNL is considered suitable and produces high stone free rates. However, high stone free rates of PCNL also seem to be accompanied by high sepsis rates. Sepsis is a life treating infection. Over the last few years, the introduction of RIRS for smaller stones (< 2) cm has reduced PCNL treatments, because of its low complications rates. We performed a comparative study using a generalized additive logistic model for all sizes renal stones between

PCNL and RIRS and made recommendations in terms of stone free and sepsis. This study will help urologists to identify which types of patients are suitable candidates for PCNL and which for RIRS.
Chapter 5: Many dynamic processes in the biomedical sciences can be fundamentally described by means of differential equations. This deductive modelling is rarely matched by empirical, inductive approaches: exploratory statistical modelling of dynamic processes have rarely taken the underlying differential equation as a starting point. In this chapter we aim to overcome this apparent gap in the statistical functional modelling literature. Our aim is to define a generalized additive model for a biomedical response variable. The method uses the linear nature of the model to derive fast explicit estimators, without the need of explicitly solving the ODE.

Bibliography

[1] R. H. Baayen, D. J. Davidson, and D. M. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412, 2008.

[2] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[3] D. Belsey, E. Kuh, and R. Welsch. Regression diagnostics: Identifying influential observations and sources of collinearity, 1980.

[4] N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25, 1993.

[5] E. W. Cheney and G. G. Lorentz. *Approximation theory III*. Academic Press, 1980.

[6] M. Clark. Generalized additive models: getting started with additive models in r. *Center for Social Research, University of North Dame, Notre Dame, IN, USA*, page 13, 2013.

[7] D. Cox. iregression models and life tablesi (with discussion). *Journal of the Royal Statistical Society, B*, page v34, 1972.

[8] A. J. Dobson and A. Barnett. *An introduction to generalized linear models*. CRC press, 2008.

[9] N. DRAPER and H. SMITH. *Applied Regression Analysis*. Wiley New York, 3rd edition, 1998.

[10] B. Engel and A. Keen. A simple approach for the analysis of generalizea linear mixed models. *Statistica neerlandica*, 48(1):1–22, 1994.

[11] J. J. Faraway. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. CRC press, 2005.

[12] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.

[13] J. H. Friedman. Multivariate adaptive regression splines. *The annals of statistics*, pages 1–67, 1991.

[14] A. Gałecki and T. Burzykowski. *Linear mixed-effects models using R: A step-by-step approach*. Springer Science & Business Media, 2013.

[15] A. R. Gilmour, R. Thompson, and B. R. Cullis. Average information reml: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, pages 1440–1450, 1995.

[16] H. Goldstein. Multilevel statistical models 2nd edition (arnold, london). 1995.

[17] P. Green and B. Silverman. Nonparametric regression and generalised linear models. *London: Chapman Hall*, 1994.

[18] A. Guisan, T. C. Edwards, and T. Hastie. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*, 157(2):89–100, 2002.

[19] T. Hastie and R. Tibshirani. Generalized additive models. *Statistical science*, pages 297–310, 1986.

[20] T. Hastie and R. Tibshirani. Generalized additive models for medical research. *Statistical Methods in Medical Research*, 4(3):187–196, 1995.

[21] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

[22] T. J. Hastie. In chambers jm and hastie, tj (eds.) statistical models in s, 1993.

[23] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Springer, 2013.

[24] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[25] K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[26] X. Lin and D. Zhang. Inference in generalized additive mixed modelsby using smoothing splines. *Journal of the royal statistical society: Series b (statistical methodology)*, 61(2):381–400, 1999.

[27] B. D. Marx and P. H. Eilers. Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, 28(2):193–209, 1998.

[28] P. McCullagh and J. A. Nelder. *Generalized linear models*, volume 37. CRC press, 1989.

[29] C. E. McCulloch. Generalized linear models. *Journal of the American Statistical Association*, 95(452):1320–1324, 2000.

[30] C. E. McCulloch and J. M. Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.

[31] J. A. Nelder and R. Baker. Generalized linear models. *Encyclopedia of Statistical Sciences*, 1972.

[32] J. Pinheiro and D. Bates. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, 2006.

[33] D. Pregibon. Logistic regression diagnostics. *The Annals of Statistics*, pages 705–724, 1981.

[34] C. R. Rao and S. K. Mitra. *Generalized inverse of matrices and its applications*, volume 7. Wiley New York, 1971.

[35] G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data*. Springer Science & Business Media, 2009.

[36] S. Wood. Stable and efficient multiple smoothing parameter estimation for generalized additive models jr stat soc ser a 99: 673–686. *Find this article online*, 2004.

[37] S. Wood. *Generalized additive models: an introduction with R*. CRC press, 2006.

[38] S. N. Wood. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):413–428, 2000.

[39] S. N. Wood. Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):95–114, 2003.

[40] S. N. Wood and N. H. Augustin. Gams with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological modelling*, 157(2):157–177, 2002.